

Multi Camera Soccer Player Tracking

*Thesis submitted in partial fulfilment
of the requirements for the award of the degree of*

Master of Technology

in

Electronics & Instrumentation

by

Dipen Chandra Mondal

Roll no - 212EC3158

Under the supervision

of

Prof. Sukadev Meher



Department of Electronics & Communication Engineering
NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA
राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला
May 2014

Multi Camera Soccer Player Tracking

*Thesis submitted in partial fulfilment
of the requirements for the award of the degree of*

Master of Technology

in

Electronics & Instrumentation

by

Dipen Chandra Mondal

Roll no - 212EC3158



Department of Electronics & Communication Engineering
NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA
राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला
May 2014

Abstract

Now a day's spread of super computers, existing of high resolution and low-priced video cameras, and increasing the computerized video analysis has made more curiosity in tracking algorithms. Automatic identification and tracing of multiple moving objects through video scene is an interesting field of computer visualization. Identification and tracking of multiple people is a vital and challenging task for many applications like human-computer interface, video communication, security application and surveillance system. Various researchers offer various algorithms but none of this was work properly to distinguish the players automatically when creating occlusion.

It is difficult to evaluate the path of an object when tracing in busy places, where a huge number of persons are continuously moving and occlude to each other in the image plane. However, this problem occurs more challenging because sometimes players and grass field may have same features, for example same team players wear same colored jersey or player's jersey color and grass field color are same. In this situation it is an extremely challenging task for a single camera view to cover entire field region to distinguish each players position and shape when one player becomes overlap partially or hidden by another player. A Common methodology is needed to install multiple cameras and collecting information from multiple camera angles and then develops the tracking scheme. In this thesis we discussed about tracking multiple players positions in soccer game by using multi-camera setup.

The first step to tracking multiple objects in video sequence is detection. Background subtraction is a very popular and effective method for foreground detection (assuming that background should be stationary). In this thesis we apply various background subtraction methods to tackle the difficulties like changing illumination condition, background clutter and camouflage. The method we propose to overcome this problem is operates the background subtraction by calculating the Mahalanobis distances.

The second step to track multiple moving objects in soccer scene by using particle filters method that estimate the non-Gaussian, non-linear state-space model, which is a multi-target tracking method. These methods are applied on real soccer video sequences and the result show that it is successfully track and distinguish the players. After tracking is done by using multi camera views, we collecting the data from all cameras and creating geometrical relationship between cameras called Homography.



Department of Electronics and Communication Engineering
National Institute of Technology Rourkela

ROURKELA-769 008, ODISHA, INDIA

May, 2014

Certificate

This is to certify that the thesis titled as "*Multi Camera Soccer Player Tracking*" by "**Dipen Chandra Mondal**" is a record of an original research work carried out under my supervision and guidance in partial fulfilment of the requirements for the award of the degree of *Master of Technology degree* in *Electronics and Communication Engineering* with specialization in *Electronics and Instrumentation* during the session 2013-2014.

Prof. Sukadev Meher

Acknowledgments

I owe deep gratitude to the ones who have contributed greatly in completion of this thesis. Foremost, I would like to express my sincere gratitude to my advisor, Prof. Sukadev Meher for providing me with a platform to work on challenging areas of Multi Camera Soccer Player Tracking. His profound insights and attention to details have been true inspirations to my research.

Next, I would like to express my respects to Prof. U. C .Pati, Prof. T. K. Dan, Prof. S. K. Patra, Prof. K. K. Mahapatra, Prof. S. K. Behera, Prof. Poonam Singh, and Prof. A. K. Sahoo for teaching me and also helping me how to learn. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I would like to thank all the faculty members and staff of the Department of Electronics and Communication Engineering, NIT Rourkela for their generous help in various ways for completion of this thesis.

I would like to thank my lab mates Mr Deepak Panda, Mr Aditya Acharya, Mr Deepak Singh, Mr Bodhisattwa Charkraborty, and Miss Sonia Das for their time to time advice and encouragement in my project work. Their help can never be penned with words. I would like to thank all my friends and especially my classmates Neelam Abhinav Karthik for all the thoughtful and mind stimulating discussions.

Most importantly, none of this would have been possible without the love and patience of my family. My family, to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heartfelt gratitude to them especially to my mother.

Dipen Chandra Mondal

Rourkela, May 2014

Contents

Abstract	i
Certificate	ii
Acknowledgement	iii
Contents	iv
List of Figures	vii
1. Introduction	1
1.1 Background and Motivation	3
1.2 Problem statement	4
1.3 Aims and Objectives.	5
1.4 System Overview	7
1.5 Outline of thesis	7
1.6 Conclusion	8
2. Background Subtraction	9
2.1 Introduction	10
2.2 Motion Detection	10
2.2.1 Frame Differencing	10
2.2.2 Optical Flow	10
2.2.3 Background Subtraction	11
2.3 Related Work.	11
2.4 Background Modelling.	12
2.4.1 Simple Background Subtraction	12
2.4.2 Motion Detection Based on Sigma-Delta Estimation ($\Sigma - \Delta$).	13
2.4.3 Effective Σ - Δ Estimation	13
2.4.4 Simple Statistical Difference (SSD)	14
2.4.5 Running Average	14
2.4.6 Gaussian mixture model (GMM)	15
2.4.7 Mahalanobis distance	15

3 Object Tracking Using Multi Cameras	17
3.1 Introduction	18
3.2 Multi-Camera with Non-Overlapping FOVs	19
3.3 Adaptive Mixture Models for Image Segmentation.	20
3.4 Establishing a Common Coordinate System	20
3.5 Camera Setup and Selecting View	21
3.6 Estimating Measurement Error Covariance	24
3.7 Fusion for Multiple Camera View	26
4 Multiple Objects Tracking	27
4.1 Introduction	28
4.1.1 Application	28
4.2 Object Representation.	29
4.3 Object Detection.	30
4.3.1 Detection by Recognition	30
4.3.2 Motion Detector	32
4.4 Tracking Methods	34
4.4.1 Morphological Operation	35
5 Particle Filter	38
5.1 Introduction.	39
5.2 Particle Filter.	39
5.3 Update Particle	41
5.4 Calculation of Log Likelihood	42
5.5 Resampling	43
5.6 Experimental Results	44
6 Homography	45
6.1 Introduction	46
6.2. Condition	47
6.3 Direct Linear Transformation (DLT) algorithm	47
6.3.1 Over-determined solution	49

4.3.2 Inhomogeneous solution	49
6.4 Projective transformation model	50
7 Conclusions and Future Work	52
8 Bibliography	53

List of Figures

Figure 1.1 Examples of computer vision used in different fields. From left to right: motion detection, sports statistics and pedestrian flow	2
Figure 1.2 Watching of sport events as technology proceeded	3
Figure 1.3 Soccer video footage	4
Figure 1.3: Block Diagram of Tracking System	6
Figure 2.1: Background Subtraction	13
Figure 3.1 Six camera angle views of a soccer game	19
Figure 3.2 an example scene and corresponding foreground image	21
Figure 3.3 (a) Showing the point correspondences footage taken from a fish-eye lens. (b) Showing measured features of the badminton court lines on the sports hall floor, used for calibration	22
Figure 3.4 Example of camera view angle	23
Figure 3.5 Chessboard shape floor image	24
Figure 3.6: Typical bounding box fit on a player	25
Figure 4.1 Object representations	30
Figure 4.2 Structuring elements	36
Figure 4.3 Erosion operations	36
Figure 4.4 Dilation operations	37
Figure 5.1 Block Diagram of Particle filter	40
Figure 5.2 Algorithm of Particle Filter	41
Figure 5.3 (a) Move Particle, (b) Resample Particle, (c) Move Particle	42

Figure 6.1 Homography	47
Figure 6.2 2D Homography	49
Figure 6.3 Projective Transformations	51

Chapter-1

Introduction

1 Introduction

Computer vision is the field of study relating to machines that observe the world around them. As a scientific it is concerned with the technology of non-natural system extracting information from images or orders of image may come from a variety of sources that may include single snap-shot cameras, video cameras or multiple synchronized cameras.

Application for computer vision increase significantly between various fields. Security is one field that uses computer vision to a great extent. Video motion detector allows for automatic impostor alerts while more advanced systems are detect suspicious people or packages in public areas. Systems that are able to measure the length of a queue or arrays of a human movement are used in shopping malls and airports to optimize staffs and layout conclusions. In a sports environment video data can be analyzed to extract statistical information such as how often are person handled the ball or which side had more possession.

This thesis considers the problem of multi camera players tracking on a sports field. This problem requires one to use techniques from several area of computer vision. Combining the various techniques in a computationally effective manner poses a challenge and reaching a real time application is not a minor problem.



Figure 1.1 Examples of computer vision used in different fields. From left to right: motion detection, sports statistics and pedestrian flow

1.1 Background and Motivation

As technology has improved over the past years, the number of people following sports around the world has increased correspondingly. Before the invention of radio the only way to gain knowledge about events from far away was by word of mouth or reading a newspaper or magazine. With the arrival of radio people can listen to live commentary as the game was unfolding. Later when television became available it became possible to watch a game happening on other side of the world. With every step forward in technology it become easier to follow sports and allowed larger audiences to follow the action.

With the greater following that sports obtained, the analysis of how players performs during matches and seasons also gained interest. All sports fans are want to know the up-to-date strategies. But when we watched the game most video footage of the players strategies are extracted from the video sequence after finish the game.

To analysis the player's performance an automatic system is required to track the player during the game. So we able to calculate the distance cover by a player and also calculate which areas of the field they spend maximum of time. This is help to develop the players performance as well as technical strategies.

The use of cameras in developing such a system will also provide various advantages above, say, attaching a GPS or other tracking device to each player. One of the big advantages is that the use of cameras is non-intrusive. Players



Figure 1.2: Watching of sport events as technology proceeded

will not need to attach a device to their clothing or person. In contact sports such tracking devices also run the risk of being damaged during physical contact.

The costs involved when using cameras also make them an attractive option. Although setup costs may be high if high- quality cameras are used, these costs are incurred only once. Using personal devices for each player would require continues costs to maintain such devices. A final advantage is that cameras are passive medium. A solution using radio or radar waves might work in a similar fashion to a camera-based solution, however it would need to project those waves onto the field. This may interface with transmission of audio and video feeds to viewers around the world.

1.2 Problem Statement

Tracking problem is defined as followed an object at time t , detect the possible location of the same object at time $t+1$ [1]. In this project players will be tracked during a short video footage captured from a football match, broadcast on a television. Figure 1.3 shows an example of the soccer video frame. Following difficulties are arising during the tracking of multiple objects.

Occlusion happens when multiple objects are overlap. During soccer game occlusion happen simultaneously, as a result players are hidden partially or full

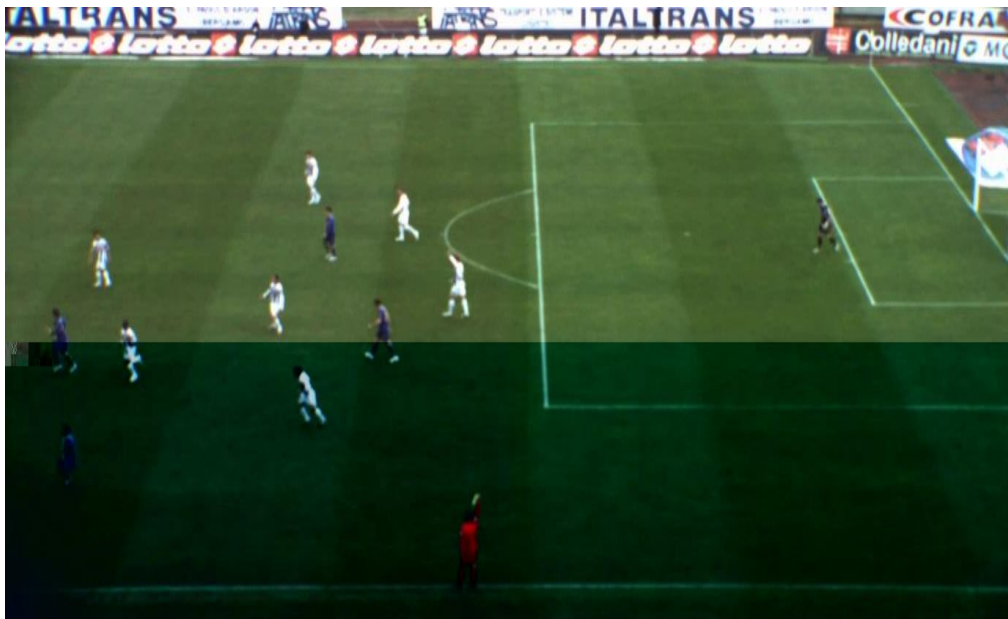


Figure 1.3 Soccer video footage

Background clutter create problem when background appear same as players. In soccer game this can happen when players jersey color are same as soccer field color or painted any sponsors logos on the field or crowded scene becomes the background when player is near the crowd.

Player motion is nature of the game. During the game, players are changing their position as well as direction consequently and also change their speed suddenly. So it is difficult to track them continuously.

Players' appearances are considered over short periods of times for many reasons including running, passing and tackle the ball or raising an arm.

Environmental changes like lighting and weather affect are creating problem the appearance of the players.

Camera Motion creates motion errors as a result we face panning and zooming problem. When players are close to the camera angle it looks bigger and brighter and when they are fur from the camera its look smaller and blur. So player detection is difficult.

1.3 Aims and Objectives

The primary goal of this project is to design and implement such a computer vision system that is able to track players as they move around on a sports field. Different components are compared based on accuracy and computational complexity to find a suitable compromise. The components are then combined into a complete system that is able to detect and track players moving around the ground, as observed by several cameras. Attempts must be made to have the system run in real time (approx. 30 frames/sec) to allow relevant statistics to be gained during the playing of a match. The work in this thesis is limited to detecting and tracking players in non-contact sports such as field hockey and soccer, due to the increased difficulties that arise from re-identifying players after complex multi person contact situations.

The setup of such an arrangement, along with the amendment of the cameras, is a vital component of this task. If this is not done properly all the data that are extracts will be incorrect. As a result the setup must be accurate and easy to implement by non-experts. It should also be easy to modify the setup at the later stage.

After the setup procedure the next stage of importance is the computation that needs to be performed on each camera stream individually. For each camera stream, the system must be able to detect the players that are within the camera's field of view and, having detected them, track those players through the video the sequence. The detection stage is important, without it the system will not be able to track any players. On the other hand new players do

not enter the field very regularly. As such the detection step must be computationally inexpensive, while still being able to detect new players within a reasonable time (around 10-12 frames). The detection stage should also be robust against false positives. The tracking of the players in each video sequence forms the largest component of the system and it requires high degree of accuracy. It is of vital importance that the track does not lose the any players it is tracking as this may have a severely negative influence on the tracking accuracy.

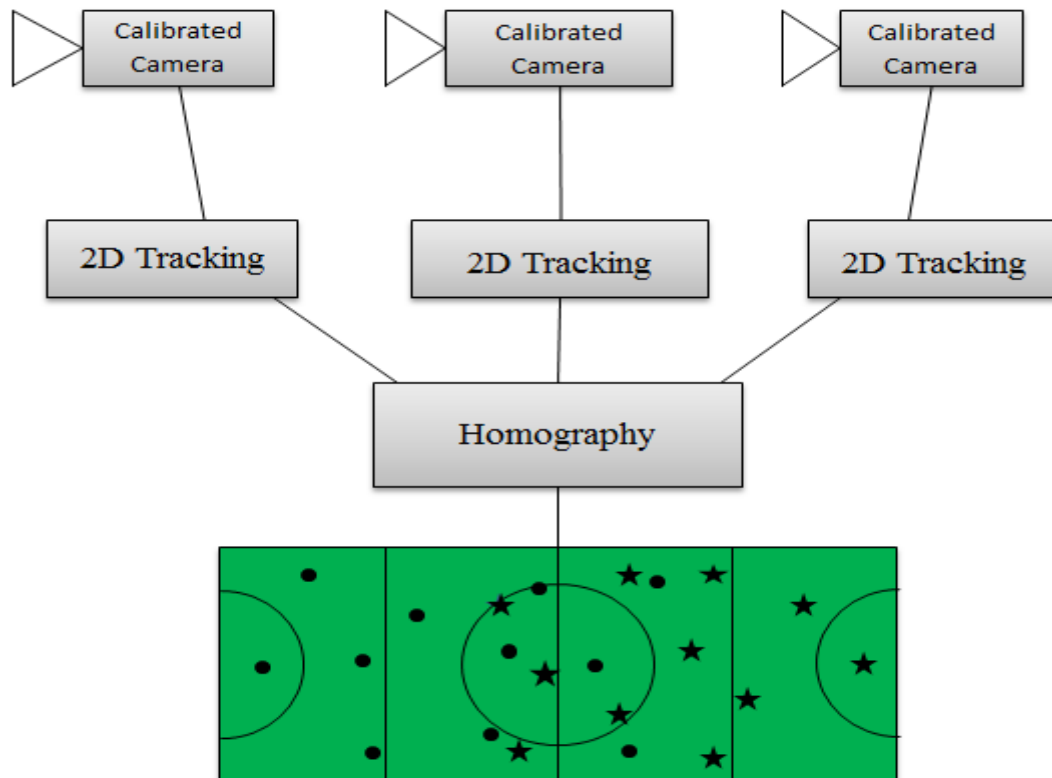


Figure 1.3: Block Diagram of Tracking System

The final component of the tracker is the 3D position estimation of players on the field. Combining the tracking data of each of the individual cameras, this must be accurately estimating the position of each player on the field. The triangulation procedure must be computationally inexpensive to increase the achievable frame-rate of the system.

An additional aim of the thesis is to implement the modules required for the various stages of the project, developing the code rather than using pre-developed modules. This allows the code that is written to be developed specifically for the given application. Writing the code oneself also gives one complete control over the code, so that modification or improvements can easily be made.

1.4 System Overview

The system is developed several component parts. This system has three core components (figure: 1.3)-

- I. Camera setup
- II. 2D tracking
- III. 3D tracking

The camera setup component is also called physical setup which is used in the field. The physical setup along with the camera calibration is needs to be calculated. Accurately setting the interior factor as well as the camera position in the world can be used to perform the measurements which it notices.

The second component is the processing which is to be performed on each video stream separately. The First stage is to detect the players who are play in the field. Motion detection is implemented to detect players who are moving around the field. After players can be detected, they are tracked through the video sequence. Finally to complete this numerous different methodologies are deliberated, and a hierarchical methodology of particle filter is executed.

The final stage is combines the tracking information from each of the individual cameras to determine the 3D position of the players on the ground. Corresponding players are found between the different view of cameras, and their position on the field is then triangulated from the multiple views.

1.5 Outline of thesis

This section details how the remainder of this thesis is organised. Chapter 2 presents a brief survey of background subtraction methods for motion segmentation along with Mahalanobis distance. In chapter 3 we have discussed about how multi cameras are setup around the football field. Object detection and tracking method is described in chapter 4 and chapter 5 respectively. Homography between six camera views is discussed in chapter 6. Finally, Chapter 7 concludes the thesis with the suggestions for future research.

1.6 Conclusion

The primary aim of this thesis is to discuss about the tracking methods, review their techniques, arrange them into different classes, and identify new fashions. In general, Object tracking is a challenging task. Difficulties arise due to abrupt object motion, changing players appearance and the scene, non-rigid object structures, object-to-object and object-to-scene overlapping, and camera motion. Generally tracking is performed in higher-level applications that required position and/or shape of the object in every frame. In a particular application assumptions are made to solve the tracking problem. In this thesis, we discuss about the multi-camera setup, various tracking methods, classification of object, how to overcome from occlusion problem and examine their pros and cons and finally we discussed the Homography.

Chapter- 2

Background Subtraction Methods

It is a fundamental and critical task to detect moving object from a video scene. There is many application to identify moving object in computer-vision system. Background subtraction is a common approach for detecting moving object in videos frame from static cameras that differs significantly from a background model. There are many difficulties to develop a good background subtraction method. First, it will be robust in illumination changes. Second, it is not to identify any non-stationary background objects such as moving leaves, rain, snow, and shadows create by moving objects. Finally, the internal background model should be changes quickly for starting and stopping the vehicles.

2.1 Introduction

Background subtraction is a vital topic in image processing and computer vision application. It is also known as Foreground Detection where foreground image is extracted for further processing. Foreground means region of interest of an object like humans, cars, text etc. It is required to know that which process is used this technique after the step of image preprocessing. Background subtraction is a common technique to detect the moving objects from the video sequence. The difference between the current frame and a reference frame is called Background image or Background model.

2.2 Motion Detection

Motion segmentation was designated into three major classes of technique as frame differencing, optical flow, and background subtraction by Hu et al [2] Motion detection targets the moving area such as players movement on the field. Detecting moving object regions provides a concentration on tracking, feature extraction and analysis. The motion segmentation is adumbrated as:

2.2.1 Frame Differencing: Frame differencing [3] makes use of the pixel-wise differences between two or three consecutive frames in an image sequence to extract moving regions. The threshold function determines change and it depends on the speed of object motion. It's hard to maintain the quality of segmentation, if the speed of the object changes significantly. Frame differencing is very adaptive to dynamic environments, but very often holes are developed inside moving bodies.

2.2.2 Optical Flow: Optical flow [4] based motion segmentation uses characteristics of flow vectors of moving objects overtime to detect moving regions in an image sequence.

Optical-flow-based methods can be used to detect independently moving objects even in the presence of camera motion. However, most of these methods are computationally complex and very sensitive to noise, and cannot be applied to in real time without specialized hardware.

2.2.3 Background Subtraction: Background subtraction is a popular method for motion segmentation, but it requires a static background. It detects moving regions from video sequences by taking the difference between the current image and the reference background image in a pixel-by-pixel manner. It is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events etc. Therefore, it is highly dependent on a good background model to reduce the influence of these changes [5], as part of environment modelling.

2.3 Related Work

The background subtraction [6], [7], [8], [9] and [10] is the most popular and common approach for motion detection. The idea is to take the difference between the current image and model image of background by using thresholding procedure. It gives silhouette region of an object. This approach is simple and computationally affordable for real-time systems, but is extremely sensitive to dynamic scene changes from lightning and extraneous event etc. Therefore it is highly dependent on a good background maintenance model. The problem with background subtraction is to automatically update the background from the incoming video frame and it should be able to overcome the following problems:

- **Motion in the background:** Non-stationary background regions, such as branches and leaves of trees, a flag waving in the wind, or flowing water, should be identified as part of the background.
- **Illumination changes:** The background model should be able to adapt, to gradual changes in illumination over a period of time.
- **Memory:** The background module should not use much resource, in terms of computing power and memory.
- **Shadows:** Shadows cast by moving object should be identified as part of the background and not foreground.

- **Camouflage:** Moving object should be detected even if pixel characteristics are similar to those of the background.
- **Bootstrapping:** The background model should be able to maintain background even in the absence of training background (absence of foreground object).

For moving object tracking proper detection is important. It's hard to get the entire above problem solved in one background subtraction technique. So the idea is to simulate different background subtraction techniques which are available in the literature and compare experimental results for different soccer videos.

2.4 Background Modelling

Background modelling is the heart of any background subtraction technique. Background model is that which robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest.

2.4.1 Simple Background Subtraction

In simple background subtraction an absolute difference is taken between every current image. $I_t(x, y)$ and the reference background image $B(x, y)$ to find out the motion detection mask $D(x, y)$. The reference background is generally the first frame of a video, without containing foreground object.

$$D(x, y) = \begin{cases} 0, & \text{if } |I_t(x, y) - B_t(x, y)| \leq \tau \\ 1, & \text{Otherwise} \end{cases}$$

Where, $I_t(x, y)$ = Current frame

$B(x, y)$ = Background frame

τ = Threshold

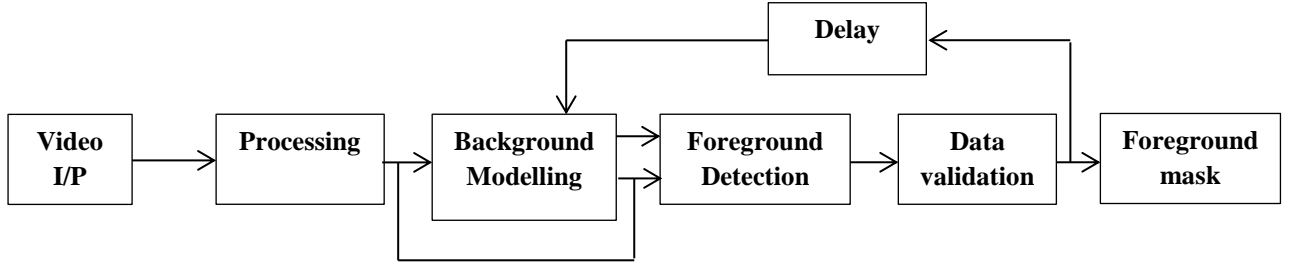


Figure 2.1: Background Subtraction

2.4.2 Motion Detection Based on Sigma-Delta Estimation ($\Sigma - \Delta$)

The $\Sigma - \Delta$ background estimation is a simple non-linear method of background subtraction [11]. It is a recursive computation of a valid background model of the scene. However, this model degrades quickly under slow or varying light conditions, due to the integration in the background model of pixel intensities belonging to the foreground objects.

2.4.3 Effective $\Sigma - \Delta$ Estimation

An $M \times N$ resolution digital image was taken where x and y are spatial coordinates and the original input image $F_f(x, y)$ is defined bellow.

$$F(x, y) = \begin{bmatrix} F_f(0,0) & F_f(0,1) \dots & F_f(0,N-1) \\ F_f(0,1) \vdots & F_f(1,1) \ddots & F_f(1,N-1) \vdots \\ F_f(M-1,0) \dots & F_f(M-1,1) & F_f(M-1,N-1) \end{bmatrix}$$

In McFarlane's $\Sigma - \Delta$ estimation algorithm [30], the new value $B_f(x, y)$ of background is determined by the previous background value $B_{f-1}(x, y)$ plus $\text{sgn}(F_f(x, y) - B_{f-1}(x, y))$. There new background values $B_f(x, y)$ do not consider the attribute of the original input image $F_f(x, y)$. Therefore when the moving objects are slowing down, stopping, or frequently appearing, the ghost effect are occurred in their built background images. In order to improve the ghost effect in the built background image, temporary input image $F_f^*(x, y)$. When $F_f^*(x, y)$ is not equal to $F_f(x, y)$ i.e., $F_f(x, y)$ belong to the attribute of moving object; the new background value $B_f(x, y)$ does not need to be adjusted in this frame. Otherwise, the new background value $B_f(x, y)$ must be adjusted with the $\Sigma - \Delta$

background estimation. Let $C(x, y)$ be the counter for each pixel at the coordinate (x, y) , α be the sampling interval of the frames and can be represented as $\text{sgn}(\alpha) = 1, \text{if } \alpha > 0, \quad \text{sgn}(\alpha) = -1, \text{if } \alpha < 0, \quad \text{and } \text{sgn}(\alpha) = 0, \text{if } \alpha = 0$ and τ be the threshold of $C(x, y)$. When $C(x, y)$ is less than or equal to τ , a new value of $(F_f^*(x, y) + B_{f-1}(x, y))$ divided by 2 is used to replace the new background $B_f(x, y)$. It can adjust the background value $B_f(x, y)$ quickly to approach the real background value. Otherwise background video is adjusted by the sgn function with a multiple interval of α .

2.4.4 Simple Statistical Difference (SSD)

This Simple Statistical Difference [SSD] technique calculates the mean and the standard deviation $\sigma_{x,y}$ for individual pixel (x, y) in the background image having K images in the time intermission (t_0, t_{k-1})

$$\mu_{x,y} = \frac{1}{K} \sum_{k=0}^{K-1} I_k(x, y)$$

$$\sigma_{x,y} = \left\{ \frac{1}{K} \sum_{k=0}^{K-1} [I_k(x, y) - \mu_{x,y}]^2 \right\}^{\frac{1}{2}}$$

In motion detection, modification between the current image $I_t(x, y)$ and the mean $\mu_{x,y}$ from the background images is designed.

$$D(x,y) = \begin{cases} 1, & \text{if } [I_t(x, y) - \mu_{x,y}] \geq \lambda \sigma_{x,y} \\ 0, & \text{otherwise} \end{cases}$$

2.4.5 Running Average

Simple background subtraction cannot handle illumination variation and results in noise in the motion detection mask. The problem of noise can be overcome, if the background is made adaptive to temporal changes and updated in every frame.

$$B_t(x, y) = (1 - \alpha)B_{t-1}(x, y) + \alpha I_t(x, y)$$

Where α is a learning rate. The motion detection mask $D(x, y)$ is calculated as follows:

$$D_{x,y} = \begin{cases} 1, & \text{if } |I_t(x, y) - B(x, y)| \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

τ = Threshold value

2.4.6 Gaussian mixture model (GMM)

The GMM methodology models each pixel history as a cluster of Gaussian type distributions and uses an on-line approximation to update its parameters. As per this method, the background is found as the expected value of the distribution corresponding to the most populated cluster [12]. The stability of the Gaussian distributions is evaluated to estimate if they are the result of a more stable background process or a short-term foreground process. Each pixel is classified to be a background if the distribution that represents it is stable above a threshold. The model can deal with lighting changes and repetitive clutter. The computational complexity is higher than standard background subtraction methods. This methodology is greatly improved on grounds of performance by considering recursive equations to adaptively update the parameters of the Gaussian model. [13]- [14]

A pixel at time t is modelled as mixture of K Gaussian [10] distributions. The probability of observing the current pixel value is given by

$$P(x_t) = \sum_{i=1}^K W_{i,t} * \eta[x_t, \mu_{i,t}, c_{i,j}]$$

Where, $W_{i,t}$, $\mu_{i,t}$ and $c_{i,j}$ and are the estimate weight, mean value and covariance matrix of I_{th} Gaussian in the mixture at time t is the Gaussian probability density function equation (2)

2.4.7 Mahalanobis distance

Proposed by Mahalanobis (1936), the Mahalanobis distance is a distance that accounts for probability distribution and equals the Euclidean distance under the standard normal distribution. It is a unit less measure.

A group of observation $x = (x_1, x_2, x_3, \dots, x_N)^T$ with mean $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^T$, then the covariance matrix S is defined as-

$$D_M = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} .$$

A dissimilarity measure can also define by Mahalanobis Distance. Dissimilarity measure of two random vectors \vec{x} and \vec{y} with same distribution with the covariance matrix S is

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} .$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidian distance. If the covariance matrix is diagonal matrix, then the resulting distance measure is called a normalized Euclidean distance.

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{s_i^2}} ,$$

Where s_i is the standard deviation of the X_i and y_i over the sample set.

Chapter- 3

Object Tracking Using Multi Cameras

From past years, it has been realized that the most difficult task to track multiple objects in soccer game is occlusion. It is a challenging task for a single camera view to cover entire field region to solve each players position and shape when one player becomes overlap partially or hidden by another player. A Common methodology is needed to install multiple cameras and collecting information from multiple camera angles and then develops the tracking scheme.

3.1 Introduction

Now a day object tracking using multiple cameras has become progressively popular, mostly for surveillance applications. These system creating the track of persons by a set of non-overlapping cameras [15], the camera handoff problem of passing ball from one video sequence to another, as the target leaves one camera view and enters another [16], using colour information to re-identify a pedestrian if he/she re-appears in the scene, or in the view of an independent neighbouring camera [17], and to aid occlusion reasoning by fusing the information from multiple views of the same scene, using Bayesian belief networks [18]. Ellis [19] has learned the topology of a random camera network, through statistical analysis of many observations of pedestrians walking through the scene(s). Stein [20], [21] has devised an excellent method for establishing a common coordinate frame between multiple video streams. Planar geometric constraints are applied to moving objects in the scene, rather than using photometric properties, which can vary between images and cameras. This is also used to align video from multiple un-synchronised cameras in time.

Multiple cameras are used more considerably for 3D view. Bowden [22] reconstructs 3D pose from a single camera view through learning a non-linear point distribution model of a human's upper body. Ong and Gong [23] extend this project to track a human body using two cameras. A hybrid 2D-3D model (outlines plus skeleton) is learned using hierarchical PCA, and a condensation tracker fits the model to each view. Gait analysis for medical purposes is the focus of Marzani et al. [24], who find that more than three cameras are needed to disambiguate all occlusions. Jennings [25] makes use of stereo range images for 3D finger tracking, as does Harville [26] for people tracking, using the depth information to create a plan view statistic. Also of note is Kanade's use of 30 cameras at Super Bowl XXXV allowing reconstruction from any viewpoint using Virtualized Reality and also of 49 cameras in 'The 3D Room' [27].

This chapter gives an overall idea on how to track multiple objects using multiple cameras with separate views.

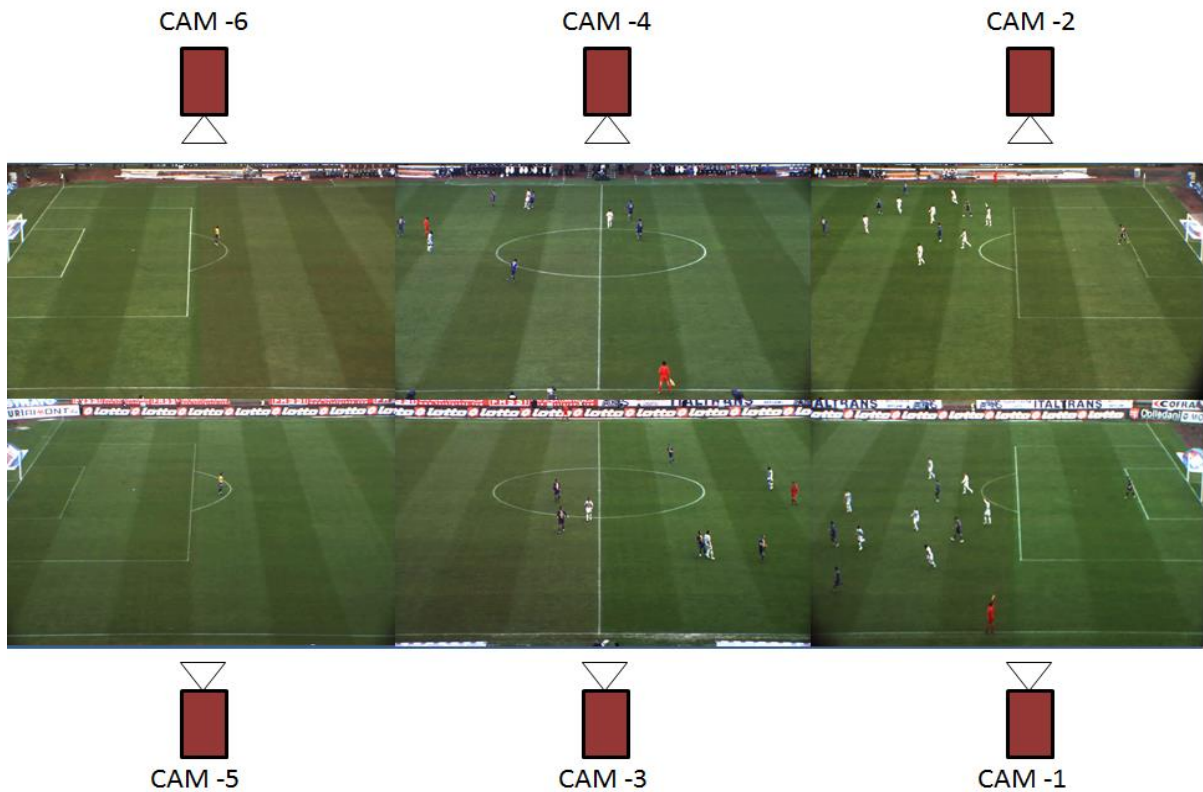


Figure 3.1 Six camera angle views of a soccer game

3.2 Multi-Camera with Non-Overlapping FOVs

Object tracking in multiple cameras with non-overlapping FoVs [28], [29] is very challenging task, as problem of correspondence occurs when an object is tracked across multiple cameras. The task in hand is to determine if the object is a new object in the scene or it is a same object that is already being tracked by some other camera. Objects are often separated in time, space, as seen from different FOVs and there is a change in how an object appears from one camera view to another. An object can take many paths across camera and generate different observations of the same object in various camera views. Because of the different placing of the cameras, it's not possible to use space-time constraints among the exists and entrance area of the camera. Here, in this we have investigated how to track objects through multiple cameras with disjoint views using object appearances in the multi - camera FOV. Object appearance can be modelled by its color or intensity, and it is a function of scene radiance, image irradiance, exposure, and camera parameters.

3.3 Adaptive Mixture Models for Image Segmentation

Over the years moving objects at a variety of speeds, shadows are cast, and lighting varies has received much attention for foreground extraction from crowded places. In [30] Stauffer and Grimson present an adaptive background mixture model. This computationally effective method is accepted here, and offers a robust foreground extraction, through the background modelling as a set of Gaussians on a per-pixel basis. Here we discussed the GMMs in image sequences.

For each pixel, a mixture of K Gaussians is formed on the RGB pixel data. A weight $W_{K,t}$ is associated with each mixture K and is updated at each time step t

$$W_{K,t} = [1 - \alpha] W_{K,t-1} + \alpha [M_{K,t}]$$

Where α =learning rate

$W_{K,t}=1$ for the mixture which matches 0 for the remaining mixtures

These weights are then re-standardised, and the mean and variance of the identical Gaussian is adaptively updated with the new observation.

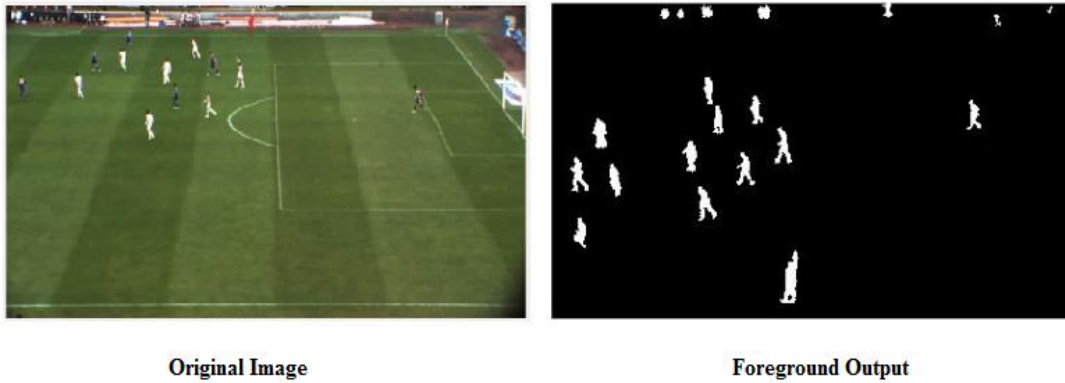


Figure 3.2 an example scene and corresponding foreground image

3.4 Establishing a Common Coordinate System

A 3D calibration is performed to improve the camera calibration. A 3D calibration allows not only points on the ground plane (x, y) to be identified on an image (u, v) , but also for three-dimensional points (x, y, z) to be identified on an image (u, v) . This now allows for

(u_{feet}, v_{feet}) and (u_{head}, v_{head}) on the image to be identified from the real world positions $(x, y, 0)$ and $(x, y, z_{player height})$ where (x, y) the ground plane position of the player. Adjusting each of the images from the different views allows points to be transformed between points in one image, points in the real world, and the corresponding points in another image of the same scene. The methods of Tsai [31] are employed to calibrate the images. These methods perform a perspective projection, using a pin-hole camera model. Many pairs of (image, world) points are needed to calibrate the image, and should be well spread over the image. As many points as are identifiable in the scene are used, typically around 50. Figure 3.3 shows an example image used for calibration, where the 51 points used in the calibration have been superimposed on top. The specific location from which the footage is taken is ideal for calibration, since there are many markings on the sports hall floor which, once measured, can easily be identified on images of the scene from any view. Without such points to use, it may be necessary to construct a large calibration frame. This convenience is not really contrived, especially in the sports domain.

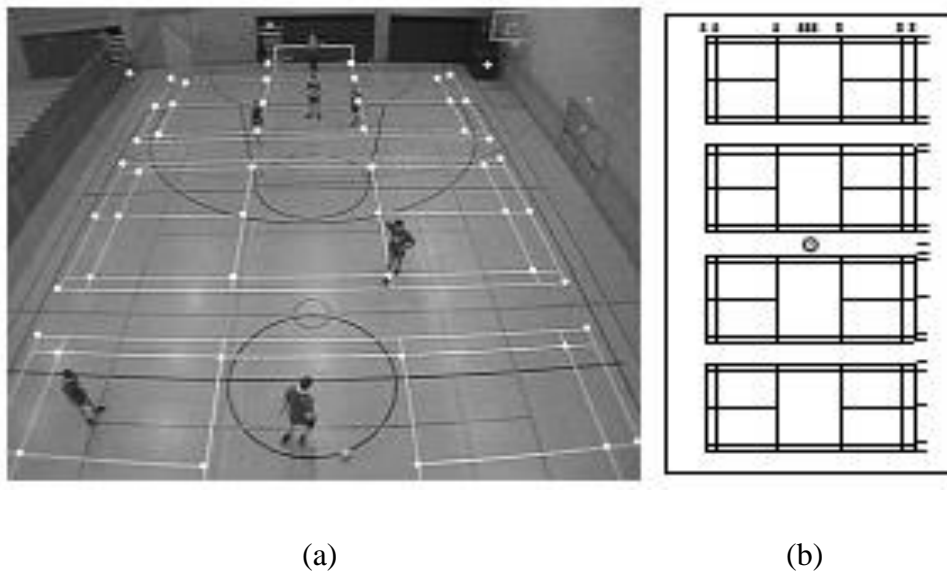


Figure 3.3 (a) Showing the point correspondences footage taken from a fish-eye lens. (b) Showing measured features of the badminton court lines on the sports hall floor, used for calibration.

3.5 Camera Setup and Selecting View

To cover large area of soccer field we use fish-eye lenses and placed the camera as high as possible. It was difficult to cover the whole of the field used a conventional camera; even it

placed several meters high of the sports gallery top. Figure 3.3 shows an example image taken from a fixed camera fitted with $\times 0.7$ fish-eye lens. The camera regulation delivers us an accurate real world to image coordinate conversion, effectively trade with the radial distortion effects which appear in the image. When set the fish-eye lens, errors were found to have a mean of 1.5 pixels, compared to 0.9 pixels with a normal camera lens. Similarly, object space errors were found with a mean of 87mm compared to 45 mm which is roughly twice as large. For that footage it's not to work well with the template matching tracking scheme as players shape changes rapidly. In tracking experiments fish-eye lens was not used for capturing the video footage because when player's shapes varied considerably camera were observed from almost above to noticed mainly heads and shoulders. In normal lens most pixels were covered the middle portion of the players which occupied very few pixels for that players looked similar in the traditional view.

Ideally, using an overhead camera facing straight down from very high above would eliminate almost all of the problems encountered. The whole pitch would be cover by a single camera, and each player (seen from above) would be symbolized by a distinct 'blob' separated from all the other players but this is impossible because the ceiling too low and no one can developed to use in outdoor situations. By choosing a camera angle similar to that used in broadcast football, there is the possibility to track successfully to cope with an outdoor 11-a-side football game, covered by multiple cameras.

There is need appropriate camera locations in rugby and netball games for different characteristics. Figure 3.4 (a) and (c) shows three different camera views which cover the whole field (perhaps three more at the other end). In rugby games it is difficult to distinguish players separately. Figure 3.4 (d) gives an example of this. In rugby games if the cameras were placed side of the field it would be hard to gives any useful information. But if the cameras were placed end of the pitch, there is more accurate separation between the players in the image and this helps for machine vision algorithms.



(a)



(b)



(c)



(d)



(e)



(f)

Figure 3.4 (a)-(c) Three views of a rugby game to cover the whole pitch. (d) Rugby viewed from the side is problematic for computer vision approaches, since the players frequently line up across the pitch, creating many instances of multiple occlusions. (e) Netballers mark each other very tightly; the players appear to move in pairs. (f) Consideration should be given to how likely the ball is to come into contact with the camera.

Netball is played in an attractively different way to soccer. But both the games are similar. Both the teams play on a restricted pitch region with a round ball with a goal at each other end and both team players are try to score as many goals as possible. However, in netball games opponents are much closer than in soccer. Figure 3.4 (e) shows each player will be next to their opposing player, so chances of occlusion are more. A final consideration is remind the placement of camera position where it is safe from being hit by a ball and also remind camera placement depends on type of games. Figure 3.4 (f) shows the ball coming close to the camera during a netball game.

3.6 Estimating Measurement Error Covariance

Kalman filtering is performed to predict the estimate positions of the players. The measurement error covariance R varies depending on the image point. Performance of filtering is improved when R changes with respect to where the player is on the ground plane. Figure 3.5 illustrates how the equal-sized chessboard squares on the ground plane appear to be of very different sizes on the image.



Figure 3.5 Chessboard shape floor image

Closer view points are more accuracy than the further away, which are at a lower resolution. For this type of view, a pixel at the front may signify 0.003m of the ground plane, whereas other side it may signify 0.5m.

On the image plane the bounding box around the player is represented by an image position (u, v) a height h , and a width w . Thus if (u_0, v_0) is the midpoint of the baseline, then

$$\begin{aligned} u_0 - \delta u < u < u_0 + \delta u \\ v_0 - \delta v < v < v_0 + \delta v \end{aligned}$$

Where $\delta u = 0.25w$ and $\delta v = 0.1h$.

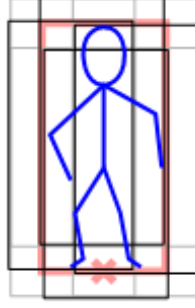


Figure 3.6: Typical bounding box fit on a player.

On the image plane the measurement error covariance matrix takes the form-

$$\mathbf{R}_{img} = \begin{bmatrix} (\delta u)^2 & 0 \\ 0 & (\delta v)^2 \end{bmatrix}$$

\mathbf{R}_{img} Varies over the image and it is depend the player position and the size of their bounding box. But this covariance matrix cannot be changed directly to the ground plane, because the non-linear image plane changes to ground plane. The ground plane measurement error covariance matrix at any point (x, y) can be change to the image coordinate (u, v) , and taking four points on the image plane $(u + \delta u, v)$, $(u - \delta u, v)$, $(u, v + \delta v)$, and $(u, v - \delta v)$ which symbolize the deviation from (u, v) . Now \mathbf{R} to be estimated as-

$$\mathbf{R} = \begin{bmatrix} \frac{1}{4} \sum_{i=1}^4 (x_i - x)^2 & \frac{1}{4} \sum_{i=1}^4 (x_i - x)(y_i - y) \\ \frac{1}{4} \sum_{i=1}^4 (x_i - x)(y_i - y) & \frac{1}{4} \sum_{i=1}^4 (y_i - y)^2 \end{bmatrix}$$

Also testing on a typical image gave-

$$\mathbf{R} = \begin{bmatrix} 144^2 & 50^2 \\ 50^2 & 194^2 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 212^2 & -65^2 \\ -65^2 & 346^2 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 226^2 & 300^2 \\ 300^2 & 700^2 \end{bmatrix}$$

(a) At the front

(b) near the middle

(c) near the back

The current tracking method includes a particle filter for tracking each player position at each time step, and estimate the next position to predict the new location of the player.

3.7 Fusion for Multiple Camera View

The main goal is to track individual player in as many views as possible, to achieve exact location of the players. Using the multiple cameras to cover the entire pitch to track players from a single view and then follow them between views.

Chapter-4

Multiple Object Tracking

Automatic moving object detection and tracking through video sequence is an interesting field of computer vision. Object detection and tracking of multiple people is a vital task for many applications like human-computer interface, video communication, security application and surveillance system.

The main goal of object tracking is tracking-by-detection and filter tracking. In the first stage, tracking relies on perfect detection of the object over various frames and the second stage is knowledge about the object's estimated movement to predict where the object is moving to next frame and make object detection easier and more precise.

4.1 Introduction

The field of computer vision, detection and tracking of object is a vital and challenging task. Tracking of multiple people before, during and after is important for many applications. However, tracing in busy places, where a huge number of persons occlude to each other, remains an interesting topic in the computer vision [32], [33], [34], [35]. Now a day's spread of super computers, existing of high resolution and low-priced video cameras, and increasing the computerized video analysis has made an excessive part of curiosity in tracking algorithms. There are three vital steps in object tracking:

- 1) Detection,
- 2) Tracking objects in frame to frame, and
- 3) Analysis the object which is being tracks to distinguish their performance.

4.1.1 Application

- Motion-based recognition (Gait representation, object detection etc.)
- Automated surveillance (Observing and identify their activities)
- Video indexing (Annotation and retrieval of the multimedia databases)
- Collaboration between human vs computer (Gesture recognition)
- Traffic monitoring (Monitoring traffic activities)
- Vehicle navigation (Roadmap, GPS system)

It is difficult to evaluation the path of an object in the image plane because objects are continuously moving around the arena. The main difficulties happening in the object tracking [36], [37] are changes in illumination, complex object shape, size and occlusion in case of tracking crowded scene. Some of the problems facing in tracking of moving objects can be shortened as follows-

- Information loss in 3-D image to 2-D plane,
- Noise in image video,
- Difficulty to recognize the exact position of moving object in every frame,
- One object become partially occlude or hidden by another object due to object or structure present in the scene,
- Complex size,
- Background motion,
- Continuous changes an illumination condition,
- Required real-time processing.

In most of the researches of object tracking, have been proposed numerous approaches. But their approaches are different to each other based on the way of application. Some questions are arising like: Which representation is suitable for object tracking? Which is used for image features? Etc. A large number of approaches have been proposed for tracking object. Stauffer and Grimson [36] have demonstrated each pixel as a mixture of Gaussians and use on-line approximations to update the model. This can give us the idea with Changes in scene lightning, background motion, and changes the long term scene. Maddalena and Petrosino [37] have modelled SOBS based on self-organization through artificial neural networks that can hold background clutter, continuing illumination changes and camouflage, no bootstrapping restrictions, overcomes the shadow problem, and detect different types of object tracking videos which is taken from still cameras. Toyoma et al. [38] can deal with the problems of illumination changing, background clutter, camouflage and shadows, and proposed three-component system for background maintenance: pixel level component, region-level component and frame-level component.

4.2 Object Representation

In object tracking, Object representation is an important task. It may be important to track object in a specific domain. Normally objects are depicted by their shape and appearances.

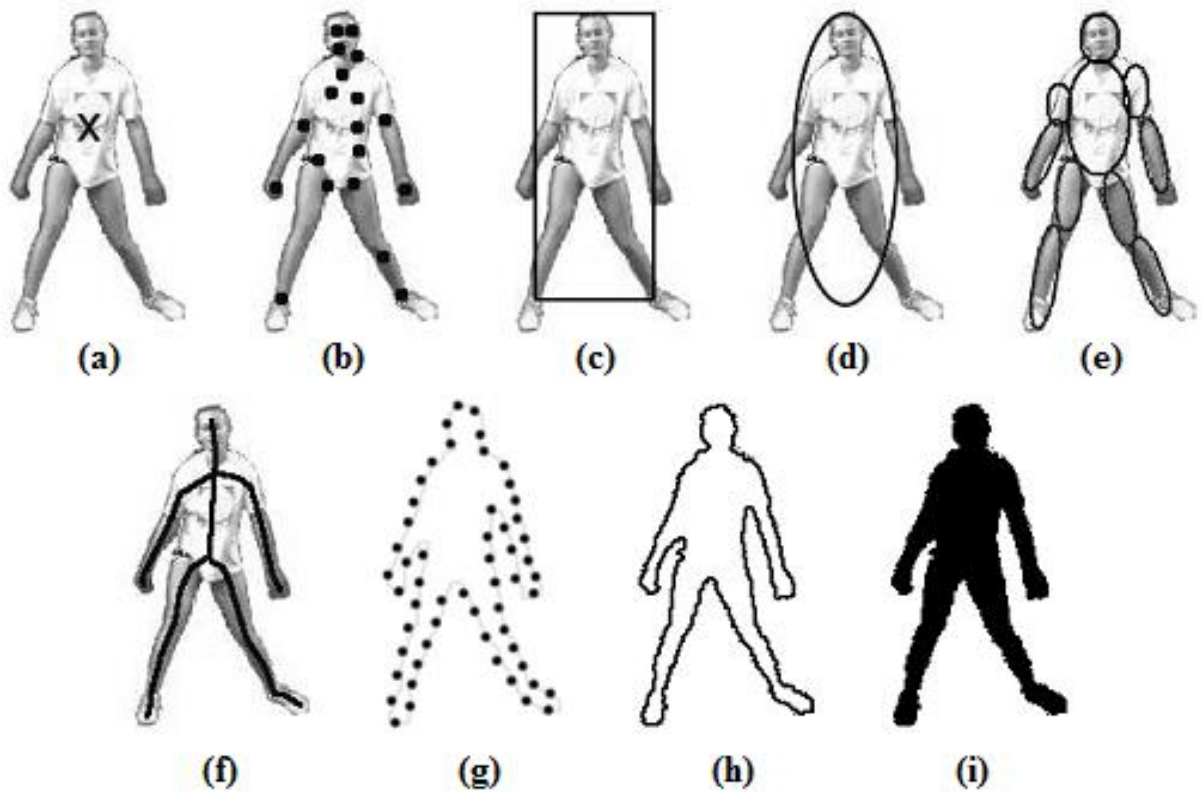


Figure 4.1 Object representations. (a) Centroid, (b) multiple points, (c) rectangular shape, (d) elliptical shape, (e) part-based multiple shapes, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette.

4.3 Object Detection

Every tracking method required an object detection mechanism to detect and locate the people in an image or video sequence. It has many real world applications, over a wide range of fields.

Two broad fields of detecting people or objects in video sequences are discussed in this section. First is to find people by trying to match some area in the image to a model of what a person looks like, known as detection by recognition. Second method is tried to find areas of motion in an otherwise static scene.

4.3.1 Detection by Recognition

Detection by recognition attempts to recognize a person or object by matching an area of the scene to some model of what a person looks like. Different algorithms use many different

features, some attempting to interpret the scene much as a person world, while other analyzes the scene using mathematical models.

- **Edge Orientation Histogram:** A popular approach used when detecting specific objects in cluttered scenes is to build some mathematical model that describes the object in a manner it can then be searched for in an image. One such approach is to model an object as a collection of edges in different directions, as in [39]. When trying to detect if a person occurs in the scene one only needs to search for an area in the scene that has a similar collection of edges. Popular edges detection method is to convolve an image with the Sobel operators [40]

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Two edge images, E_x and E_y are composed, one for each of the Sobel operators. For each of the pixel in the original image a gradient and direction can be calculated as-

$$M(x,y) = \sqrt{E_x(x,y)^2 + E_y(x,y)^2},$$

$$G(x,y) = \arctan \frac{E_x(x,y)}{E_y(x,y)} \times \frac{180}{\pi}$$

- **Scale Invariant Feature Transform:** Lowe [41] has proposed a method of object recognition by an object as a collection of interest points in a geometric relation, known as the Scale Invariant Feature Transform (SIFT). The object is then found in an image by searching for a collection of similar interest points in a similar configuration.

To locate interest points a different of Gaussian function is applied to the image sampling levels. At the first sampling level the image is smoothed using two passes of a 1D Gaussian function-

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-x^2 / (2\sigma^2))$$

in the vertical and horizontal directions.

At of the interest points a SIFT key is generated by calculating the pixel edge magnitude, $M(x, y)$, and orientation, $R(x, y)$, of the base image as-

$$M(x, y) = \sqrt{(A(x, y) - A(x+1, y))^2 + (A(x, y) - A(x, y+1))^2},$$

$$R(x, y) = \arctan \frac{A(x, y) - A(x+1, y)}{A(x, y+1) - A(x, y)}$$

Objects can now be detected in cluttered images.

4.3.2 Motion Detector

Motion detection is the process of detecting and locating areas of motion in a video sequence. Basic motion detection algorithms might only detect if there is any motion in video sequence, while more advanced algorithms try to find where in each frame the motion occurs. These algorithms typically attempts to split each frame into two areas-foreground and background. Foreground regions are those areas where motion occurs while background regions are static areas.

- **Mixture of Gaussians:** Mixture of Gaussian (MoG) is popular in many machine learning and pattern recognition algorithms as a method to build pdfs. The MoG pdfs

are constructed as a linear combination of Gaussian probability function, allowing for multi-modal behaviour that a single Gaussian world not permits.

When applied to background subtraction problems, the MoG technique attempts model regions background by a mixture of K Gaussian distributions. Each Gaussian corresponds to an aspect of interest of the pixel, such as intensity or brightness.

- **Optical Flow:** Optical flow is the process of estimating a 2D motion field between frames sequence, originally developed by Horn et al [42]. If the field can be calculated accurately motion can be found by searching for areas where there is a great deal of flow in one direction. Areas with no flow are generally background areas while areas with near constant non-zero flow indicate moving objects. Areas with flow in haphazard direction may also indicate background areas, such as foliage moving in the wind.

In optical flow methods to find a match for each pixel p in image I_t with a pixel in image I_{t+1} . Search areas are limited to surrounding the original pixel in order to reduce computational expense and increase accuracy. If such a match can be found the flow of the pixel is the displacement (d_x, d_y) of the pixel between the two images.

In algorithms, calculate the displacement as the movement in x and y directions-

$$(d_x, d_y) = (I_t(p_x) - I_{t+1}(p_x)), (I_t(p_y) - I_{t+1}(p_y))$$

Another method is to calculate the flow between images by matching a window of size $w \times h$ in image I_t with a window of the same size in image I_{t+1} . The best such match can be found by minimizing the function E -

$$E(d_x, d_y) = \sum_{x=p_x-w}^{p_x+w} \sum_{y=p_y-h}^{p_y+h} (I_t(x, y) - I_{t+1}(x + d_x, y + d_y))^2$$

A popular optical flow method was developed by Lucas et al [43] that has been used in tracking applications.

4.4 Tracking Methods

Tracking systems are implemented for many reasons containing security and surveillance system application, video communication, road traffic monitoring and hand gesture recognition and so on.

General object detection algorithm has four stages-

(1) Initialize the target object representation

For the image at time t

(2) Generate a set of candidate regions

(3) Measure the likelihood that a candidate region represents the target object

(4) Determine the most likely location of the target object. $t = t+1$

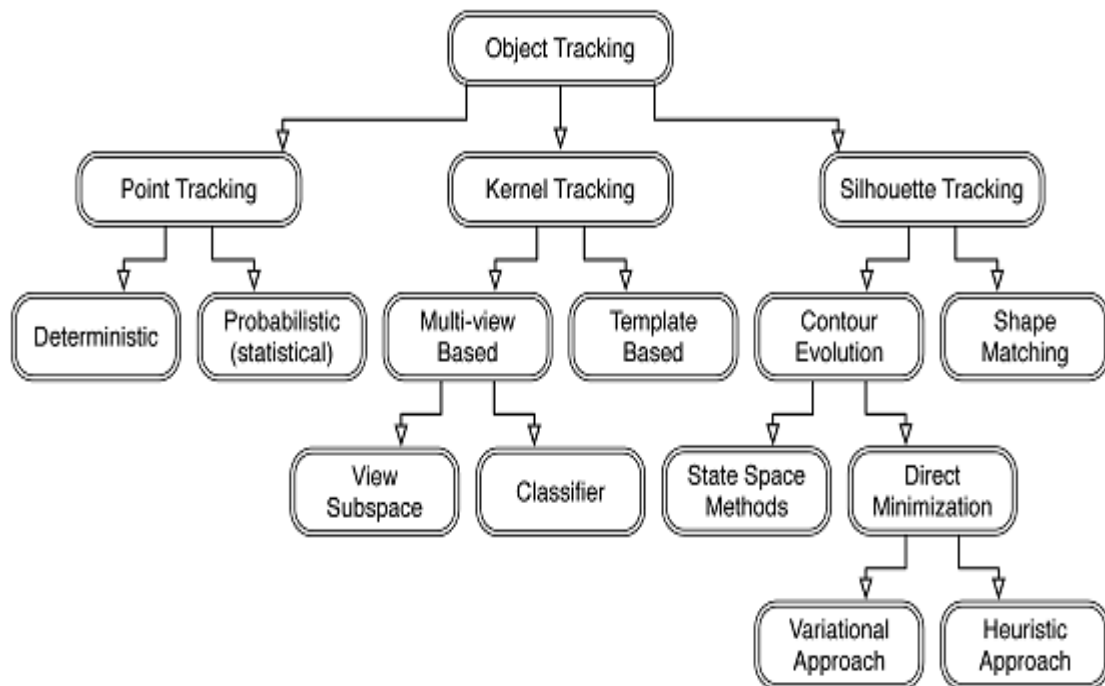
Target object: Target object is defined as the object which is being tracked. There is a method to track an object in which separates it from other objects and regions.

Candidate region: Candidate region is a region, which represents a target object at $t+1$. The technique of forming the position of candidate regions is often mentioned to as the Prediction Model.

Prediction Model: Prediction Model is guesses the potential location of the target object. Each candidate region is compared to the target object and measure the similarity. Tracking algorithms calculate this similarity using a Measurement Model.

Measurement Model: The measurement model compared to the candidates region representation and the target object representation to allow the algorithm to resolution which candidate best matches the target object.

Nomenclature of Tracking Methods:



4.4.1 Morphological Operation

Morphological image processing is used to extract image components for representation and description of region shape, such as boundaries, skeletons, and the convex hull.

Morphological operation [44] analysis an image with a small shape or template called a **structuring element**. The structuring element is placed at all probable places in the image and it is related with the equivalent neighbourhood of pixels. Some operations test whether the element "fits" within the neighbourhood, while others test whether it "hits" or intersects the neighbourhood:

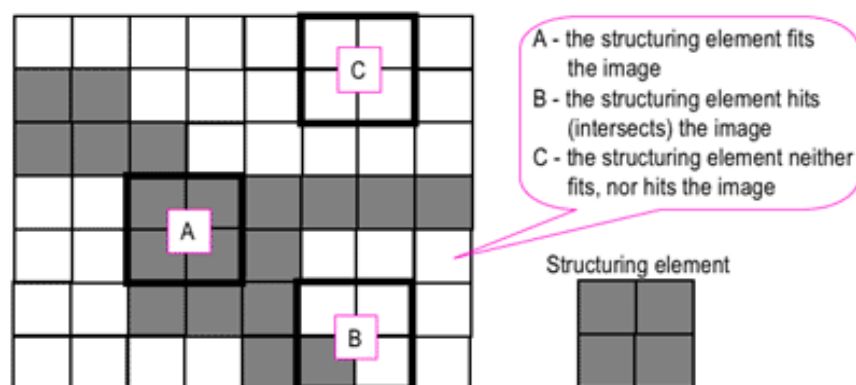


Fig 4.2 Structuring element

- **Erosion:** Erosion is a set A by structuring element B: all z in A such that B is in A when origin of B=z

$$A \ominus B = \{z | (B)_z \subseteq A\}$$

- Erosion means shrink the object

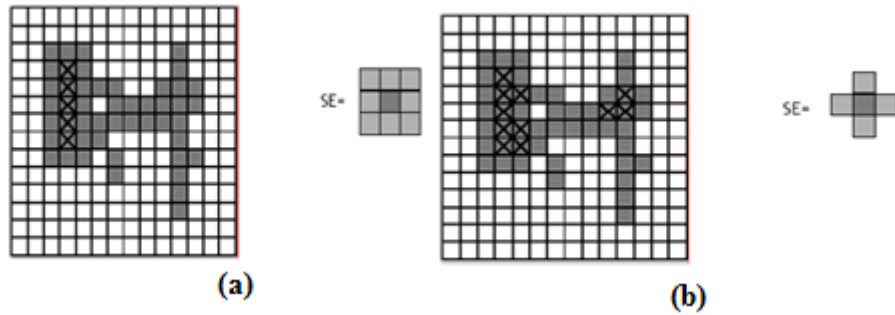


Figure 4.3 Erosion operations

- **Dilation:** Dilation is a set A by structuring element B: all z in A such that B hits A when origin of B=z

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \Phi\}$$

- Dilation means expand the object

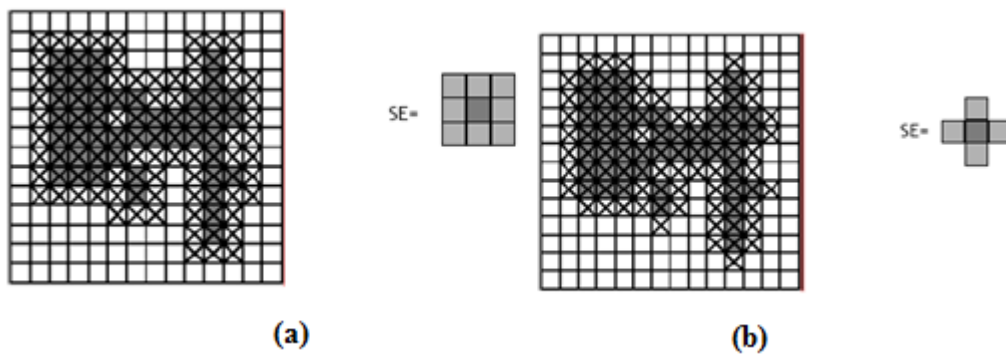


Figure 4.4 Dilation operations

- **Opening:** Opening generally smooths the contour of an object and eliminate thin protrusions.
The opening of a set A by structuring element B is defined as-

$$A \circ B = (A \ominus B) \oplus B$$

Therefore, the opening A by B is the erosion of A by B, followed by a dilation of the result by B.

- **Closing:** Closing also tends to smooth sections of contours but fusing narrow breaks and long, thin gulfs and eliminating small holes and filling gaps in the contour. Similarly, the closing of a set A by structuring element B is defined as-

$$A \bullet B = (A \oplus B) \ominus B$$

Therefore, the closing of A by B is the dilation of A by B, followed by the erosion of the result by B.

Chapter-5

Particle Filter

This thesis examines and evaluates the video scene which is taken from multi cameras for estimate the path of soccer players. It is necessary to track players for investigate the players position, path, speed and strategies for future game plan. This information studies by coach and take decision how to improve his team players techniques. Many players tracking researchers have to give many suggestions for tackle this problem but still no solution work properly to distinguish the player. Presently using particle filter is focusing to track target object.

5.1 Introduction

Automatic moving object detection and tracking through video sequence is an attractive field of computer vision system. For many years it has been paying attention of researchers how to track moving objects. When trackers track the object problem comes due to rapidly change its shape, position, speed and also direction at a short time interval. An optimum solution is Bayesian filter that uses the Bayes theorem to calculate the PDF (posterior density function). But it can't determine analytically. To solve this problem uses Kalman filter or extended Kalman filter [43]. Kalman filter solve this difficulties but it assumes that the process is linear. But when the systems or measurements are non-linear Kalman filter is failed. This time implemented the extended Kalman filter which does not require linear systems. In some cases its give unsatisfactory output. So recently developed unscented Kalman filter [44], which is based on Kalman filter and dealing with non-linearity. This technique overcome the non-linearity problem by uses the mean and covariance matrix.

5.2 Particle Filter

The main complications of tracking soccer players to track more than one player at a time. All the above methods are nearest neighbour (NN) method. Recently another method is developed called sequential Monte Carlo method which is known as condensation method or particle filtering [45]. This is based on recursive Bayesian filter by Monte Carlo simulations. The core idea is to estimate the pdf (posterior probability density function), which is estimate the track of object by a set of arbitrary particles with the help of their connected weights. There is no required any assumptions to deal with non-linearity and non-Gaussian distribution. That means higher the probability, denser the particles are concerted.

The basic block diagram of particle filter is shown below-

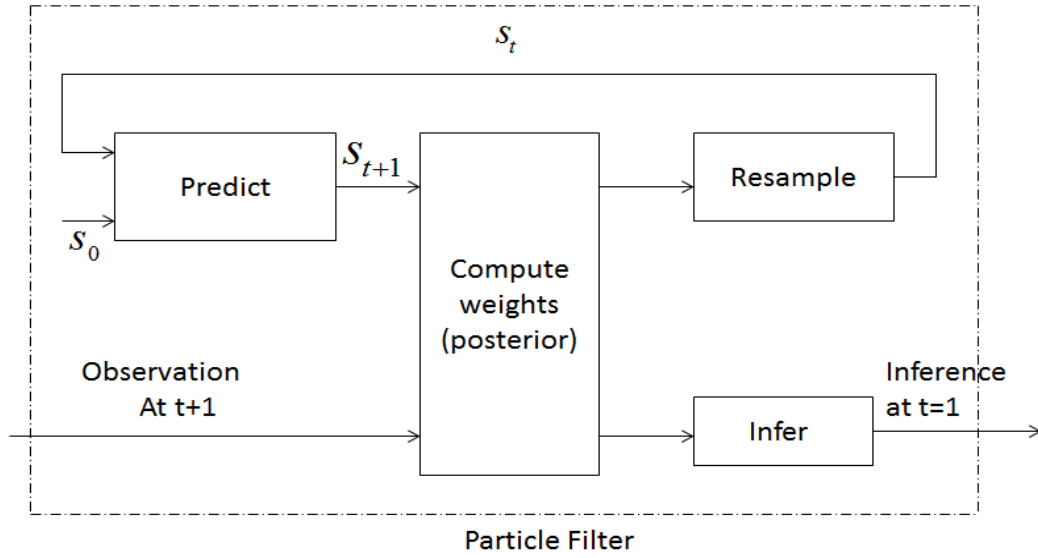


Figure 5.1 Block Diagram of Particle filter

In above Figure 5.1 system state is change at time $s_t = f(s_{t-1}, w_t)$, where s_t = system state at time t , f = state transition model and w_t = process/system dynamics, so the system is called Markovian. The system state s_t depends on the previous state s_{t-1} and the process model w_t , which is enabling in time. The system also noticed the noisy sensors $z_t = h(s_t, v_t)$. where h = observation model, v_t = sensor noise and z_t = sensor observation. Here observation model h have to made relationship between the system state s_t , sensor noise v_t and sensor observation z_t . But w_t and v_t is assumed to be known and captured by pdfs [46].

The key steps in particle Filter algorithm is shown below-

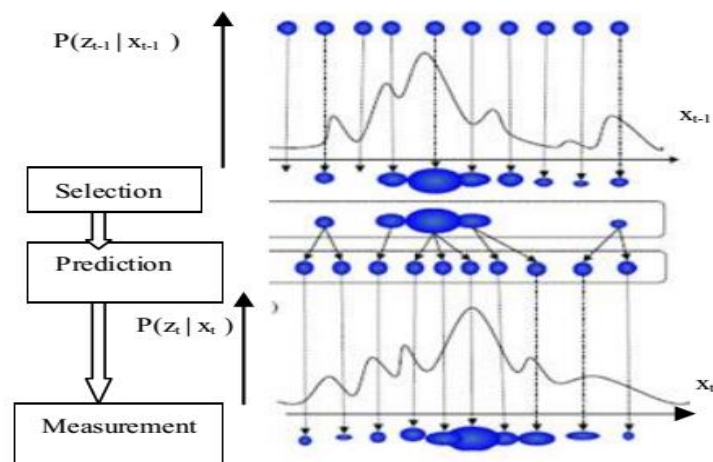


Figure 5.2 Algorithm of Particle Filter

Algorithm:

➤ At n=1

1. Sample $X_1^i \sim q(x_1 | y_1)$

2. Compute weights

$$w_1(x_1^i) = \frac{\mu(x_1^i)g(y_1|x_1^i)}{q(x_1^i|y_1)} \quad W_1^i$$

3. Resample $\{W_1^i, X_1^i\}$ to obtain N equal weights particles $\{1/N, \bar{X}_1^i\}$

➤ For n ≥ 2

4. Sample $X_n^i \sim q(x_n | y_n, X_{1:n-1}^i)$

5. Compute weights

$$\alpha_n(X_{1:n}^i) = \frac{g(y_n | X_n^i) f(X_n^i | X_{n-1}^i)}{q(X_n^i | y_n, X_{n-1}^i)} \quad W_n^i$$

6. Resample $\{W_n^i, X_{1:n}^i\}$ to find equal weight particles $\{1/N, \bar{X}_{1:n}^i\}$

➤ At each time step, we obtain estimate distributions:

$$\hat{p}(x_{1:n} | y_{1:n}) = \sum_{i=1}^N W_n^i \delta_{X_{1:n}^i}(x_{1:n})$$
$$\hat{p}(y_n | y_{1:n-1}) = \sum_{i=1}^N W_{n-1}^i \alpha_n(x_{n-1:n}^i)$$

5.3 Update Particle

For group of target tracking it is most important to study the prediction step very well for spatial illustration of uncertainties. When the process dynamics and sensor model are linear, it can be easily demonstrate by ellipsoids, spheres and polytope families.

An estimate of the variable of interest is obtained by the weighted sum of the particles. Two major stages can be distinguished- prediction and update. During prediction, each particle is modified according to the state evolution model, including the addition of random noise. In the measurement update stage, the weight is re-evaluated using a likelihood term. The residual resampling [47] is added to prevent sample poverty.

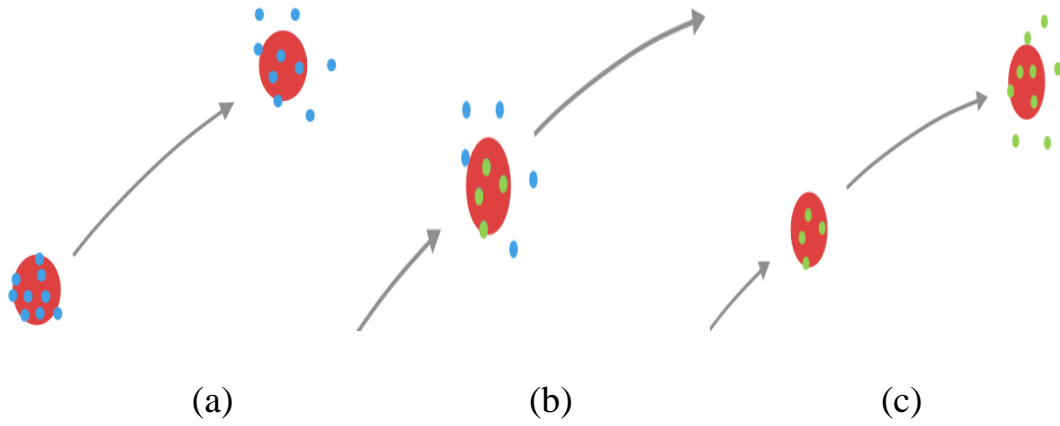


Figure 5.3 (a) Move Particle, (b) Resample Particle, (c) Move Particle

5.4 Calculation of Log Likelihood

Definition of log likelihood function is to be the natural logarithm of the likelihood function,

$$P_d = \frac{1}{\sqrt{2\pi\sigma}} * \exp\left(-\frac{d^2}{2\sigma^2}\right), \quad \text{Where, } d = \sqrt{(r-255)^2 + g^2 + b^2}$$

p_d = Probability distribution function

For many reasons, researchers often work with the loglikelihood rather than the likelihood, that is $l(\theta, x) = \log L(\theta, x)$. First reason is that $l(\theta, x)$ tends to be a simple function than $L(\theta, x)$ because in logarithmic, products are changed to summations. If $X = (X_1, X_2, \dots, X_n)$ is sample of probability distribution $f(\theta, x)$, then the likelihood is the product of X_i -

$$\begin{aligned} L(\theta; x) &= f(x; \theta) \\ &= \prod_{i=1}^n f(x_i; \theta) \\ &= \prod_{i=1}^n L(\theta; x_i) \end{aligned}$$

Loglikelihood is the sum of individual loglikelihood-

$$\begin{aligned}
l(\theta; x) &= \log f(x; \theta) \\
&= \log \prod_{i=1}^n f(x_i; \theta) \\
&= \sum_{i=1}^n \log f(x_i; \theta) \\
&= \sum_{i=1}^n l(\theta; x_i).
\end{aligned}$$

5.5 Resampling

The core idea of resampling is remove the small weight particle and generate larger weight new particles. In Particle filter a common problem is introduce that is after few iteration most of the weight of particle except few become zero or missing. So to prevent it, an efficient resampling method [48] is used. This resampling process is done by creating a set particle with equal weights $1/N$ and removes the small weight.

Algorithm:

1. Let $c_1 = 0$
2. For $i = 2 : N$

$$c_i = c_{i-1} + w_k^i$$

End For
3. $\mu_1 \sim U[0, N^{-1}]$ (uniform distribution)
4. For $j = 1 : N$

$$u_j = u_1 + N^{-1}(j-1)$$

While $\mu_j > c_i$

$$i = i + 1$$

End while

$$x_k^{j*} = x_k^j$$

$$w_k^j = 1/N$$

$$i^j = i$$

End For

5.6 Experimental Results



Frame no- 8



Frame no- 20



Frame no- 27



Frame no- 37



Frame no- 53



Frame no- 62



Frame no- 106



Frame no- 145



Frame no- 169



Frame no- 192



Frame no- 240



Frame no- 260

Chapter-6

Homography

Track multiple moving objects in soccer scene by using particle filters method that estimate the non-Gaussian, non-linear state-space model, which is a multi-target tracking method. These methods are applied on real soccer video sequences and the result show that it is successfully track and distinguish the players. After tracking is done by using multi camera views, we collecting the data from all cameras and creating geometrical relationship between cameras called Homography.

6.1 Introduction

In this chapter we discussed about the planer transformations. Consider taking a point $X=(u,v,1)$ in one image and $X'=(u',v',1)$ in another one image from slightly different camera angle. Then depict both the two images in a planer scene. So mathematically the relationship between these two images can be described as Planar Homography or Projective Transformation.

In Figure 6.1, on the plane S , the relationship between the point X_1 and the point X_2 on the camera images S_1 and S_2 , at the point X in 3D space from 2 cameras C_1 and C_2 create a 3 by 3 matrix H is called Homography. It is figured between more than 4 parallel sets of points between 2 camera images, and gives one-to-one correspondence of points.

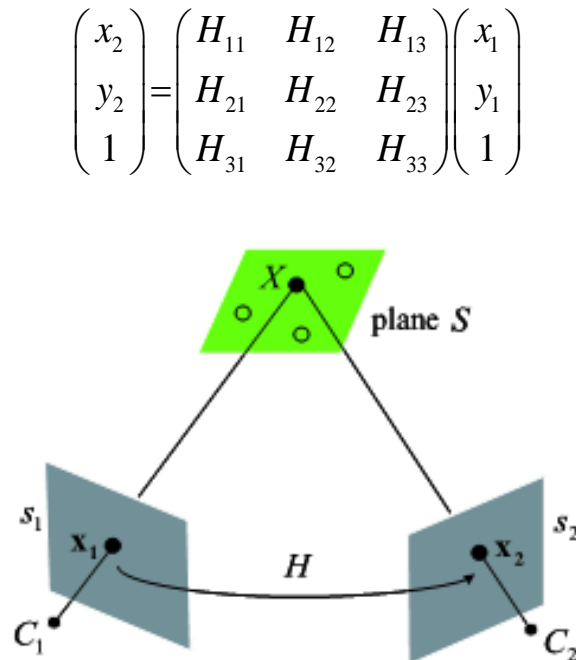


Figure 6.1 Homography

6.2. Condition

Two images are related by a Homography if and only if both images are viewing the same plane from a different angle and both images are taken from the same camera but from a different angle. Camera is rotated about its centre of projection without any translation. Note that the Homography relationship is independent of the scene structure. It does not depend on what the cameras are looking at and what is seen in the images.

6.3 Direct Linear Transformation (DLT) algorithm

In Direct Linear Transformation (DLT) algorithm, we determine the Homography matrix H has a set of four 2D to 2D point correspondences, $x_i \leftrightarrow x'_i$. The transformation equation is $x'_i = Hx_i$. In this equation 3 vectors x'_i and Hx_i are not equal, their direction is the same but magnitude is different. So the equation is expressed as $x'_i \times Hx_i = 0$ (vector cross product). The derivation of H is shown below-

Let h_j^T be the j -th row of H , so we may write-

$$Hx_i = \begin{pmatrix} h^{1T} x_i \\ h^{2T} x_i \\ h^{3T} x_i \end{pmatrix} \quad 6.1$$

We know that $x'_i \times Hx_i = 0$ then

$$x'_i \times Hx_i = \begin{pmatrix} y'_i h^{3T} x_i - w'_i h^{2T} x_i \\ w'_i h^{1T} x_i - x'_i h^{3T} x_i \\ x'_i h^{2T} x_i - y'_i h^{1T} x_i \end{pmatrix} = \mathbf{0} \quad 6.2$$

Since

$$\begin{bmatrix} \mathbf{0}^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & \mathbf{0}^T & -x'_i x_i^T \\ -y'_i x_i^T & x'_i x_i^T & \mathbf{0}^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = \mathbf{0}. \quad 6.3$$

Where $A_i h = 0$, where A_i is a 3×9 matrix, and h is a 9-vector, so the matrix H equal to-

$$h = \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix}, \quad H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \quad 6.4$$

Each pair of points generates two equations of the form

$$\begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{x}_i^T & y'_i \mathbf{x}_i^T \\ w'_i \mathbf{x}_i^T & \mathbf{0}^T & -x'_i \mathbf{x}_i^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0 \quad 6.5$$

This will be written as- $A_i h = 0$, where A_i is now the 2×9 matrix. If there are n pairs we get $2n$ equations on 9 unknowns. In our task 4 points are enough.

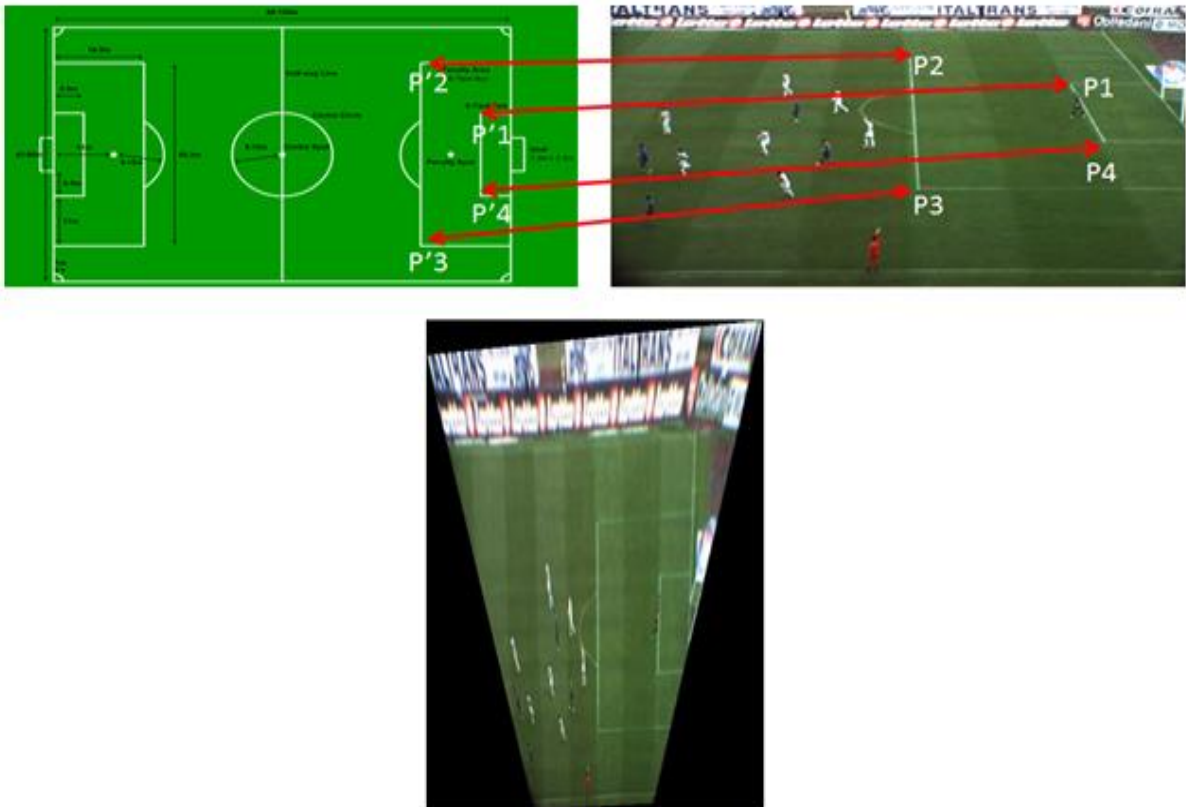


Figure 6.2 2D Homography

6.3.1 Over-determined solution

In equation no. 6.5, if we take more than four points correspondences, $x_i \leftrightarrow x'_i$, then $A_i h = 0$ is called over determined. The location of the points are exactly same as matrix A still have rank 8 if there is precise solution of h and it is to found in one dimension null space. There is no precise solution of the over determined process $Ah=0$ instead of zero solution. Nothing to exact solution, we take approximate solution with help of vector h that minimize the cost function. But now the question is what should be minimized? Answer is $h \neq 0$. Generally we used the norm like as $\|h\| = 1$ which is not important. So there is no exact explanation of $ah=0$, only to minimize the norm $\|ah\|$. There is a same problem to minimize the quotient $\|ah\|/\|h\|$. The final answer is to take a least eigen value of $A^T A$. so the algorithm is called DLT algorithm.

6.3.2 Inhomogeneous solution

To solve the homogeneous vector h, we convert it to be an inhomogeneous vector as $h_j = 1$. The solution is chosen up to the scale. We have chosen the scale so that $h_j = 1$ by observing the solution. The resulting equations derived from (6.5) are given

$$\begin{bmatrix} 0 & 0 & 0 & -x_i w'_i & -y_i w'_i & -w_i w'_i & x_i y'_i & y_i y'_i \\ x_i w'_i & y_i w'_i & w_i w'_i & 0 & 0 & 0 & -x_i x'_i & -y_i x'_i \end{bmatrix} \tilde{h} = \begin{pmatrix} -w_i y'_i \\ w_i x'_i \end{pmatrix}$$

If $h_j = 1$, where h_j corresponds to H_{33} and \tilde{h} is containing 8 vector with 8 components. After all we get a matrix equation of the form $M\tilde{h} = b$, where M=8 columns and b=8 vector. This equation solve by another techniques where M consist of 8 rows.

Objective

Given $n \geq 4$ 2D to 2D point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the 2D homography matrix \mathbf{H} such that $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$.

Algorithm

- (i) For each correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ compute the matrix \mathbf{A}_i from (4.1). Only the first two rows need be used in general.
- (ii) Assemble the $n \times 9$ matrices \mathbf{A}_i into a single $2n \times 9$ matrix \mathbf{A} .
- (iii) Obtain the SVD of \mathbf{A} (section A4.4(p585)). The unit singular vector corresponding to the smallest singular value is the solution \mathbf{h} . Specifically, if $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ with \mathbf{D} diagonal with positive diagonal entries, arranged in descending order down the diagonal, then \mathbf{h} is the last column of \mathbf{V} .
- (iv) The matrix \mathbf{H} is determined from \mathbf{h} as in (4.2).

The basic DLT for \mathbf{H}

6.4 Projective transformation model

In this section we examine the camera view point when capture images from arbitrary angle.

The relationship between the point $\mathbf{w} = [u, v, 0]^T$ and the position $\mathbf{X} = [x, y]^T$ is below

$$\begin{aligned} \lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \begin{bmatrix} \phi_x & \gamma & \delta_x \\ 0 & \phi_y & \phi_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \tau_x \\ \omega_{21} & \omega_{22} & \omega_{23} & \tau_y \\ \omega_{31} & \omega_{32} & \omega_{33} & \tau_z \end{bmatrix} \begin{bmatrix} u \\ v \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \phi_x & \gamma & \delta_x \\ 0 & \phi_y & \phi_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{12} & \tau_x \\ \omega_{21} & \omega_{22} & \tau_y \\ \omega_{31} & \omega_{32} & \tau_z \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \end{aligned} \quad 6.6$$

Multiply the two 3×3 matrices, we get-

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad 6.7$$

This is known as Projective Transformation or Homography.

The Homography in Cartesian coordinates form is written as-

$$\begin{aligned} x &= \frac{\phi_{11}u + \phi_{12}u + \phi_{13}u}{\phi_{31}u + \phi_{32}u + \phi_{33}u} \\ y &= \frac{\phi_{21}u + \phi_{22}u + \phi_{23}u}{\phi_{31}u + \phi_{32}u + \phi_{33}u} \end{aligned} \quad 6.8$$

In short,

$$X = \text{hom}[w, \phi]$$

6.9

The Homography can map any four points in the plane to any other four points.

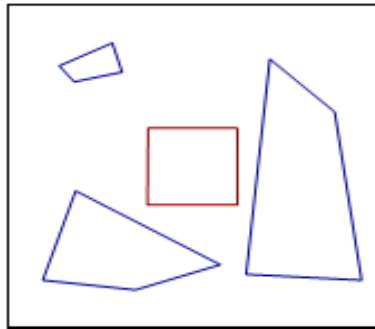


Figure 6.3 Projective Transformations

Chapter-7

7 Conclusions and Future Work

In this thesis the problem of estimating the player positions on field was discussed. Past techniques, essential constituents and overall system design were all covered in some detail.

At the start of the thesis several existing approaches to position estimation of people in a known environment were discussed in methodology and with regards to their strengths and weakness. This allowed us to identify possible routes to follow and areas to focus on.

Motion detection is simple and effective player detection solution and particle filter was found to be a useful tool for 2D tracking. The result obtained for the system were very promising overall. While some improvements can be made the system is able to solve the initial problem to a satisfactory extent. The lesser goal of real time processing was not achieved during implementation, but the use of graphical processing unit along with some algorithm optimization should achieve this goal with relative ease.

Future work should reduce the sensitivity of the algorithm and may increase the accuracy, as well as the usability of the system.

Bibliography

- [1] Interpreting Human Gesture with Computer Vision - Coutaz, Crowley (1995)
- [2] Weiming Hu, Tieniu Tan, Liang Wang, and S. Maybank. A Survey on Visual Surveillance of Object Motion and Behaviours. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 34(3):334–352, August 2004.
- [3] Alan J. Lipton, Hironobu Fujiyoshi, and Raju S. Patil. Moving target classification and tracking from real-time video. In Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98), WACV '98, pages 8–. IEEE Computer Society, 1998.
- [4] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, Performance of optical flow techniques. INTERNATIONAL JOURNAL OF COMPUTER VISION, 12:43–77, 1994.
- [5] S. Toral, M. Vargas, F. Barrero and M. G. Ortega. “Improved sigma-delta background Estimation for vehicle detection.” Electronics letters. vol. 45, no. 1 pp. 32-34, 2009
- [6] O. Barnich and M. Van Droogenbroeck. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. IEEE Transactions on Image Processing, 20(6):1709–1724, June 2011.
- [7] Ahmed Elgammal, David Harwood, and Larry Davis. Non-parametric model for background subtraction. In FRAME-RATE WORKSHOP, IEEE, pages 751–767, 2000.
- [8] K Toyama, J Krumm, B Brumitt, and B Meyers. Wallflower: principles and practice of background maintenance. Proceedings of the Seventh IEEE International Conference on Computer Vision, 1(c):255–261, 1999.
- [9] L. Maddalena and A. Petrosino. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Transactions on Image Processing, 17(7):1168–1177, 2008.
- [10] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:809– 830, 2000.
- [11] Chris Stauffer, W. Eric, and W. Eric L. Grimson. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:747–757, 2000.
- [12] A. Manzanera and J. C. Richefeu, “A new motion detection algorithm based on $\Sigma - \Delta$ background estimation,” Pattern Recognition Letters., vol. 28, no. 3, 2007.
- [13] Edited by Hanna Goszczynska. Object Tracking. INTECH open access, February, 2011.

- [14] Omar Javed, Khurram Shafique, Zeeshan Rasheed, and Mubarak Shah. Modeling inters camera space-time and appearance relationships for tracking across non-overlapping views. *Computer. Vis. Image Underst.*, 109(2):146–162, February 2008.
- [15] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real time tracking. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 246–252, Fort Collins, Colorado, 1999
- [16] S. L. Dockstader and A. M. Tekalp. Multiple camera fusion for multi-object tracking. In *Proc. IEEE Workshop on Multi-Object Tracking*, pages 95–102, 2001.
- [17] T. J. Ellis, D. Makris, and J. K. Black. Learning a multiple-camera topology. In *Proc. Joint IEEE Intl. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003.
- [18] O. Javed, S. Khan, Z. Rasheed, and M. Shah. Camera handoff: Tracking in multiple uncalibrated stationary cameras. In *Proc. IEEE Workshop on Human Motion*, pages 113–118, 2000.
- [19] V. Kettner and R. Zabih. Bayesian multi-camera surveillance. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 253–259, Fort Collins, Colorado, 1999
- [20] L. Lee, R. Romano, and G. Stein. Monitoring activities from multiple video streams: Establishing a common coordinate frame. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):758–767, 2000.
- [21] J. Orwell, P. Remagnino, and G. A. Jones. Multi-camera colour tracking. In *Proc. IEEE Intl. Workshop on Visual Surveillance*, pages 53–68, Fort Collins, Colorado, 1999.
- [22] G. P. Stein. Tracking from multiple view points: Self-calibration of space and time. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1521–1527, Fort Collins, Colorado, 1999.
- [23] R. Bowden, T. A. Mitchell and M. Sarhadi. Reconstructing 3D pose and motion from a single camera view. In *Proc. British Machine Vision Conference*, pages 904–913, Southampton, UK, 1998.
- [24] C. Jennings. Robust finger tracking with multiple cameras. In *Proc. Intl. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*.
- [25] F. Marzani, E. Calais, and L. Legrand. A 3-D marker-free system for the analysis of movement disabilities - an application to the legs. *IEEE Trans. on Information Technology in Biomedicine*, 5(1):18–26, 2001.
- [26] E-J. Ong and S. Gong. Tracking hybrid 2D-3D human models from multiple views. In *Proc. IEEE Workshop on Modelling People*, Corfu, Greece, 1999.
- [27] M. Harville. Stereo person tracking with adaptive plan-view statistical templates. In *Proc. ECCV Workshop on Statistical Methods in Video Processing*, pages 67–72, Copenhagen, Denmark, 2002.
- [28] H. Saito, S. Baba, M. Kimura, S. Vedula, and T. Kanade. Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3D room. In *Proc.*

Second Intl. Conf. on 3-D Digital Imaging and Modeling, pages 516–525, Ottawa, Canada, 1999.

[29] P. Figueroa, N. Leite, and R. Barros. Tracking soccer players using the graph representation. In *ICPR2004: Proceedings of the 7th International Conference on Pattern Recognition*, pages IV: 787–790, 2004.

[30] L. Fuentes and S. Velastin. Tracking-based event detection for CCTV systems. *Pattern Analysis Application*, 7(4):356– 363, 2005.

[31] S. McKenna, S.Jabri, Z.Duric, A. Rosenfield, and H. Wechsler. Tracking group of people. *Computer Vision and Image Understanding*, 80:42–56, 2000.

[32] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. In *CVPR2004: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages II: 406–413, 2004.

[33] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Comput. Surv.* 38(4):13, December 2006.

[34] Deepak Kumar Panda and Sukadev Meher. Robust real-time object tracking under varying illumination condition. In *INDICON 2011 - Proceedings: 2011 IEEE INDICON*, pages Birla Institute of Technology and Science, Pilani; IEEE; IEEE India Council; IEEE Hyderabad Section, 2011.

[35] Chris Stauffer, W. Eric, and W. Eric L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:747–757, 2000.

[36] K Toyama, J Krumm, B Brumitt, and B Meyers. Wallflower: principles and practice of background maintenance. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1(c):255–261, 1999.

[37] L. Maddalena and A. Petrosino. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE Transactions on Image Processing*, 17(7):1168–1177, 2008.

[38] K. Engel, “Real-Time Volume Graphics”, A. K. Peters Ltd, 2006

[39] W. Freeman, M. Roth, “Orientation histograms for hand gesture recognition”, *International Workshop on Automatic Face and Gesture Recognition*, pp. 296-301, 1994

[40] D. Lowe, “Object Recognition from local scale-invariant feature”, *International Conference on Computer Vision*, 1150-1157, 1999

- [41] B. Horn, B. Schunck, "Determining optical flow", AI Memo 572, Massachusetts Institute of Technology, 1980
- [42] D. Lucas, T. Kanade, "An interactive image registration technique with an application to stereo vision", Imaging Understanding Workshop, pp. 121-130, 1981
- [43] Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing (3rd Edition). Prentice Hall, 3rd editions, August 2007
- [44] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking," IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, Feb. 2002.
- [45] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to linear system," Proc. of Aerosense: The 11th International Symposium on Aerospace/Defense sensing, Simulation and Controls, Orlando, Florida, pp. 182-193, 1997.
- [46] G. Welch and G. Bishop, An introduction to Kalman filter, TR 95-041, University of North Carolina at Chapel Hill, 2001
- [47] D. Forte and A. Srivastava, —Resource-Aware Architectures for Particle Filter Based Visual Target Tracking,|| 2011.
- [48] E. Wan and R. van der Merwe, The Unscented Kalman Filter, Ch. 7: Kalman Filtering and Neural Networks. Edited by S. Haykin, pp. 221–280, Wiley Publishing, September 2001.